DOCUMENT RESUME

ED 084 281 TM 003 289

AUTHOR	Hiller, Jack H.; And Others
TITLE	On the Uses of the Computer for Content Analysis in Educational Research.
PUB DATE	Feb 73
NOTE	21p.; Revised version of paper presented at national conference of Association for Computing Machinery (San Francisco, August 1969)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Audiovisual Aids; Classification; *Computer Programs; *Content Analysis; Educational Research; *Measurement Instruments; *Scoring; Social Sciences; *Structural Analysis; Technical Reports

ABSTRACT

Current efforts to take advantage of the special virtues of the computer as an aid in text analysis are described. Verbal constructs, category construction, and contingency analysis are discussed and illustrated. Mechanical techniques for reducing human labor when studying large quantities of verbal data have been sought at an increasing rate by researchers in the behavioral sciences. Whatever the purpose of research, if it is to have a scientific character, ti must involve an attempt to reduce natural language data, by formal rules, to measures reflecting theoretically relevant properties of the text, its source, or its audience effects. At the present time, there is no one theory or method dominating the field of natural language analysis. Although much work is currently being expended to implement a finite set of rules on the computer, little has been accomplished that is directly useful to researchers in the social sciences. (Author/CK)

ED 084281

TM 003 289

Feb 1973

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

On the Uses of the Computer For Content Analysis in Educational Research[1]

Jack H. Hiller, Southern Illinois University

Gerald A. Fisher, Illinois Institute of Technology, Chicago

Donald R. Marcotte, Wayne State University, Detroit

FILMED FROM BEST AVAILABLE COPY

On the Uses of the Computer

For Content Analysis in Educational Research.

Jack H. Hiller, Southern Illinois University
Gerald A. Fisher, Illinois Institute of Technology, Chicago
Donald R. Marcotte, Wayne State University, Detroit

Abstract

Mechanical techniques for reducing human labor when studying large quantities of verbal data have been sought at an increasing rate by researchers in the behavioral sciences. Research interests may range from attempts to simulate human judgemental performance, as when the computer is programmed to grade student compositions, to efforts to construct and test theories of verbal behavior, to exhaustive searches after empirical relationships in textual data. Whatever the purpose of the research, if it is to have a scientific character, then it must involve an attempt to reduce natural language data, by formal rules, to measures reflecting theoretically relevant properties of the text, its source, or its audience effects.

At the present time, there is no one theory or method dominating the field of natural language analysis. Different questions seem to lend themselves to different analytic techniques. The elegant mathematical models of Chomsky and others carry with them the promise that a natural language is built and employed in accordance with a finite set of rules. Although much work is currently being expended to implement such rules on the computer, little has been accomplished that is directly useful to researchers in the social sciences.

In this paper, the authors describe current efforts to take advantage of the special virtues of the computer as an aid in text analysis. In particular, verbal constructs, category construction, and contingency analysis are discussed and illustrated with recent investigations.

The underlying basis of all current methods of computer assisted content analysis is the process of direct comparison. Words, or letters, sentences, or punctuation marks preselected by the researcher are simply matched against text data to determine occurrence rates for these units (Stone, Dunphy, Smith, and Ogilvie, 1966), or to discover the contextual associates of the pre-selected units (Laffal, 1965). Alternatively, an internal comparison of textual units (e.g., words or idioms) not previously specified by the researcher may be performed to compile word or concept frequency tables (Sedelow and Sedelow, 1968; Carlson, 1966), or to compile word cluster lists by the use of factor analysis (Iker and Harway, 1969).

Two requirements are set for the units of analysis: they must be capable of being recognized by a suitably programmed computer, and they must also have relevance to theory or the goal of research. Let us stress that theory is a necessary guide to the selection of the units of analysis, and that theory, whether profound or trivial, is yet the work of the human. The computer in all such analyses acts merely as a clerk or calculator, but never as theorist. However, this description of the computer's role is not universally agreed to.

Page and Paulus (1968), for example, had a computer programmed to count commas, periods, and classes of words, such as common words, to enable computer grading of essays. Based on the frequency statistics of such units in a student's essay, the computer printed out one of several previously stored phrases. In describing this computer activity, Page and Paulus (1968, page 162) asserted, "the computer begins to understand what it is told by the student and is able to intelligently respond to him." To the contrary, the intelligence manifest by such non-adaptive computer programs reflects the understanding not of the computer, but that of the theory reflected in the computer program.

In an explicitly stated theoretical approach to essay grading research, qualities of writing described by stylists were translated into exemplifying words and phrases, which the computer was programmed to identify (Hiller, Marcotte, and Martin, 1969). The frequency of these units in a set of essays was then measured by the computer and found to bear statistically significant correlations with the writing quality grades independently assigned the essays by expert human graders. The research philosophy found here is, we believe, common to most content analyses performed by computer. We, therefore, will discuss this research and its rationale in detail.

NATURE OF TEXT ANALYSIS

Any text forming the natural language data to be studied is exactly what it is. To rephrase its meaning, to reduce it to kernel sentences, or to measure it in some way or ways cannot in principle provide the researcher with a truly equivalent representation of the original text. The very nature of analysis is the process of abstracting from the corporate whole characteristics of the text which are defined by theory. Again, the nature of such characteristics and the techniques employed to measure their appearance in text derive from theory. In attending only to certain characteristics, others must necessarily lie unnoticed. Thus, analysis serves to reduce a text from its unity, on the one hand, and from its potentially infinite reserve of characteristics, on the other, to a restricted set of measures. We may say the text is reduced to its theoretically important essentials by analysis and measurement.

The text characteristics measured have two uses: description and inference. Description is achieved by having the computer count units of direct interest; for example, the number of sentences written, their average length, and the number of words in the text, or the frequency of specific words themselves of direct

interest, such as the word "the," or the punctuation "," comma. The other use is directed to measurement of properties of text capable of leading to an inference concerning the source of the text, or its potential effects on some audience, or, in fact, its informational content. Operationally, descriptive and inferential measures may be identical. For example, suppose we are testing the notion that long sentences (those over 20 words) typically produce comprehension difficulties. Until this hypothesis is substantiated, long sentences would tentatively represent the inferred property--difficult style. Now assume that we accept the hypothesis and are training students to write with an emphasis on reading ease. Feedback to the student would be a description of his sentence lengths.

STRATEGIES OF ANALYSIS

Determination of a text's meaning by computer is currently impossible and will continue to be unmanageable even if parsers capable of automatically reducing sentences to kernels are perfected. Assume, for example, that we wish to determine if a given text contains a certain message. An obvious fact of language is that utterances which bear the same sense with respect to some point of view may yet have different graphemic expression. "These are certainly the facts of life," and "These certainly are the facts of life," will for most purposes be equivalent. Yet, if one sentence were the message to be sought in the text, and the other were in the text, a computer programmed to perform a simple direct word by word match of message to text would fail to discover the message. A suitable parser would convert both message and text to kernel sentences which then might be effectively compared. However, it is apparent that messages comprized of more than a single sentence might not be effectively analyzed, even with the perfected sentence parser. Let us illustrate this point with two passages, each formed from identically worded sentences.

C'EST LA VIE

One day by chance, Mary happened to encounter John. Mary and John had a love affair. Lovely were the buds of Spring, and lovelier yet Summer's blooms. But then, as in all life's adventures, Summer turned to Fall, and Fall decayed into Winter's death. Mary endured, despite the finality. On one light Spring day, fickle Mary met Peter, and promptly married him. Despite her initial anxieties, she lived happily forever after.

THE ADULTRESS

One light Spring day, fickle Mary met Peter, and promptly married him. But then, as in all life's adventures, Summer turned to Fall, and Fall decayed into Winter's death. Mary endured despite the finality. One day by chance, Mary happened to encounter John. Lovely were the buds of Spring, and lovelier yet the blooms of Summer. Mary and John had a love affair. Despite her initial anxieties, she lived happily forever after.

It may be seen here that sentence order provides a contribution so crucial to the meaning of these passages that a simple comparison of their sentences or kernels would falsely determine equivalence. This conclusion of false equivalence might be refuted by requiring that matched sentences occur in identical order. However, any such restriction could also lead to an invalid result. For example, the meaning of the second passage would not importantly be affected if the order of the sentences, "Mary and John had a love affair," and "Lovely were the buds of Spring ..." were to be interchanged. It is certain that analyses performed only at the level of the sentence, with all sentences treated in isolation, will often lead to inaccurate text analysis.

Another approach to content analysis divorces itself from syntax. In this approach, the researcher limits the computer's task to a search for discrete

textual units which have been preselected to enable measurement of specified text characteristics. As stated earlier, such characteristics have two distinguishable uses; namely, description and inference.

A descriptive analysis is performed simply by having the computer identify and count text units. In general, the analyst's work in determining appropriate units for a given description is simple. A description of the "number of words" contained in a text clearly requires that words be selected as the unit of measure. (A minor problem arises, however, concerning treatment of hyphenated words and idiomatic expressions or phrases which offer a single sense through use of two or more written words. For example, in "John is a used car salesman," we may wish to treat "used car," as a single idea and hence as a single word). In the analysis, all written words are treated as an equivalence class--All Words--and all such equivalence classes are termed "categories."

CATEGORIES

A category is a list of items (e.g., words, letters, punctuation, etc.) selected by the researcher to represent a text property. The items selected for inclusion in a category are chosen on the assumption that all possess a common feature. Thus, category items may be highly dissimilar along a number of dimensions, provided all possess the attribute defining the category. In addition to the category All Words, we may also define Sentence Words and have the computer count the number of words in each text sentence, and in addition provide statistical descriptions, such as average sentence length or standard deviation of sentence length Another category could be four letter words, five letter words, and so on. These particular examples of word categories do not reflect any notion of sense or meaning.

Perhaps the most common type of meaningful category is the synonym list.

Another category type, which relates to meaning but does not rely on synonimyty, is the special purpose category, of which there are infinitely many possible. For example, we may define the special purpose category--color words--which would contain the items: aqua, blue, green, etc., or the category--cars introduced in America during 1945--and so on Such categories as provided in the above examples offer the researcher no particular difficulty regarding the selection of items for category inclusion. The concepts underlying these categories rather directly imply their items.

CONSTRUCTS

However, certain concepts of potential interest do not immediately suggest the category items. A concept or construct is an abstraction which attains meaning within a theory. As an abstraction, a construct cannot itself be directly observed. But any interesting construct must permit prediction of observable behaviors related to it. Anxiety for example, is a construct which psychologists have employed to explain aspects of speech. Mahl (1959), hypothesizing that speech disturbances could reflect anxiety, correlated counts of speech disturbances of patients with subjective listener ratings of their anxiety, and found the predicted relationship. Mahl, in his study of speech disturbances, ~~for example,~~ encountered a validation problem. He noted that his listener ratings may have been contaminated by the raters' knowledge of the speech disturbance hypothesis. Because of this, the test of the hypothesis lacked validity. In general, validation of the index designed to represent a construct constitutes the most important and difficult phase of research.

Techniques for validating the instrument of measurement are illustrated in the following section by describing research in which validation of a particular construct, vagueness, and its hypothesized category representation (operational definition) were pursued.

RESEARCH ON VAGUENESS: AN EXAMPLE OF CONSTRUCT VALIDATION

Vagueness has been defined as a "psychological construct which refers to the state of mind of a communicator who does not sufficiently command the facts, knowledge or understanding required for maximally effective communication," (Hiller, 1969). Vagueness is an internal stimulus condition which develops in a speaker or writer as he commits himself to deliver information he can't remember or simply doesn't know. It is assumed that the experienced communicator has learned a set of verbal responses which enable him to move on from his point of difficulty. Based on observations of verbal behavior, vagueness response categories were formulated and their items selected (Hiller, 1969). For example, one of the categories was termed "approximation" and among the items chosen as clues were: almost, about as, kinda, sorta, pretty much, etc. The complete set of categories is presented in figure 1.

It was hypothesized that writers who are vague in thinking would as a consequence, write less effectively. A set of 250 student essays was processed by computer to measure the vagueness represented by each essay, and these measures were then correlated with essay grades provided by human judges who knew nothing of the vagueness construct. Vagueness was found to correlate -.26 with scores for essay Content, and -.32 with scores for Creativity (both correlations are significant at $p < .0005$; Hiller, Marcotte, and Martin, 1969).

Insert Figure 1. about here

In a second test of the construct, teachers' lectures were studied to investigate the relationship between lecturing effectiveness and vagueness. A measure of the teacher's lecturing effectiveness was obtained by administering a multiple choice test of lesson comprehension to the teacher's class immediately after he had presented a 15 minute lecture. The complete set of teacher lectures

which contained over 100,000 words, was transcribed from video tape recordings and then key-punched (Gage, Belgard, Dell, Hiller, Rosenshine, and Unruh, 1971). In one set of 32 lectures, Vagueness correlated -.59 ($p < .005$) with class test scores, and in a second set of the lectures -.48 ($p < .05$, Hiller, Fisher, Kaess, 1969).

In a subsequent experiment, knowledge of speakers was manipulated to determine if vagueness, as measured through use of the vagueness response dictionary (a dictionary is a set of categories) is indeed related to the speaker's command of his subject matter. One group of speakers listened to a tape recorded lesson for 15 minutes before presenting their own lectures based on the lesson just heard. A second group listened to the lesson after 50 percent of the tape was randomly replaced with fragments of a second lesson. The group with the mutilated tape displayed a greater use of the vagueness items, $F(1,20) = 15$, $p < .001$, as had been predicted (Hiller, 1969). A sample of a vague lecture is shown in Exhibit 1.

The experiment specifically inquired if items hypothesized to be clues to vagueness could be demonstrated to occur when a speaker is working from inadequate knowledge. However, the experiment, as such, does not and could not prove the theory true. Consider the chain of inference between experiment and theory: a speaker committed to address an audience on an informative topic discovers he does not know or understand the material which has to be communicated; this inadequacy arouses an internal stimulus condition; finally, responses previously performed and reinforced to this stimulus are evoked in the speaker. The internal stimulus condition--vagueness--is a hypothetical construct whose presence must be inferred; other explanations for the experimental results may always be found.

IMPROVING THE INSTRUMENT OF MEASUREMENT

Although the measures of vagueness based on the items included in the vagueness dictionary have related to validation criteria as had been predicted, the

validity and reliability of each of the individual items in the dictionary or categories must yet be dealth with. Any given item may on certain occasions, in certain language contexts, provide an erroneous indication. In addition, certain items may be associated with intense vagueness, that is, a communicator's being completely befuddled, or with only slight uncertainty. Other of the items may quite simply have no validity as clues to vagueness. Since the vagueness dictionary contains 353 words and phrases, research to investigate all of the individual items could be overwhelming. However, given a suitable validation criteria, and given a text in which each item to be tested occurs with sufficient frequency, one manageable strategy is to split the text into two sections, one to be used to conduct a pilot study, and the other to be used to cross validate the results of the pilot study.

Selection of the items to be tested is a task at least as important as eventual validation tests. The theorist may himself select the items, or he may instruct a panel of judges to provide them. Different forms of instruction are possible. For example, the judges may be told of the construct; that is, the construct definition may be presented and explained as the basis for having each judge volunteer category items. Judges may also be presented samples of verbal behavior to have them search for response items. Or the judges, after having had the construct explained to them, may be presented a list of items tentatively selected by the researcher. The judges may accept or reject items, or they may rate items for relevance or degree of construct representation in a manner similar to the semantic differential. The researcher may employ such ratings as item weights when calculating construct representation in his data (see Holste, 1969). Items with low ratings might also be discarded from the computer search list to save computer time. However, any such exclusion of weak items may jeopardize the validity of a study. If a given writer under study happens to prefer use of an

item not included in the category, but the item happens to signify an important property of the text, then analysis without it may yield invalid results.

We must also urge that computer counts not be accepted at face value, but that any result be checked by human inspection of the text under analysis. Most of the common words of our language cloak different meanings in similar spellings (the homograph problem); thus a simple match of category words against text may lead to invalid scoring. Stone (1969) reports that a set of procedures for resolving this ambiguity through use of context clues is in preparation. But until such disambiguation procedures are available, a simple expedient is to have the computer print out each item found in text along with its context, to enable human checking. It should also be noted that use of phrases rather than single words greatly improves scoring accuracy. For example, the word "kind" may refer to "classification" or to "thoughtfulness," but it also forms part of the phrase, "kind of" or "kinda". The word "kind" in the sense of classification happens to be a vagueness item. While use of "kind" leads to many scoring errors, "kind of" is error free. Word combinations fix meaning perhaps surprisingly well.

WORD CO-OCCURENCES AS CLUES TO MEANING

In the preceeding section, it was suggested that phrases or co-occuring words may greatly improve the machine's accuracy over single words in scoring content. An interesting test of this proposal was conducted by comparing the two methods in a product simulation of the human content scoring of short-answer identification essays (Marcotte 1969). As part of a final examination in history, students were required to respond to 12 terms such a Cluniac Movement, craft guild, Cicero, etc., in a few sentences that were to demonstrate their familiarity with the terms. The course professor provided his graduate assistant an answer key for

each of the 12 terms. For purposes of research, these keys were then used by judges to grade the students' responses. The keys were also used as the basis for computer grading, as explained below.

Each key was first inspected to insure that it had simple sentences or phrases; compound and complex sentences whenever found were transformed into simple sentences. Simple sentences provided the primary computer scoring units. To economize computer time, function words (the, a, of, etc.) were eliminated from the keys. The keys were then inspected word by word to determine where synonyms were needed, and with the aid of several synonym dictionaries, each word was replaced by a category of synonyms with one or more words and phrases. Thus, each of the simple sentences of a key was recast as a set of one or more categories. Similarly, the student essays were processed by computer so that all words were either replaced by category markers or eliminated; the computer processing was also set to retain sentence units during the transformation of text to categories. Basically, the computer next matched the categories of each key with the categories of the student's response.

To be more specific, several scoring procedures were applied. In the simplest procedure, sentence organization in both the key and response was ignored and the machine simply obtained the percentage of key categories found in the student's response (the single word frequency method). In contrast to this procedure, each sentence unit of the key was matched against each unit of the response; credit was given the response only if all categories in the key sentence were found in any one of the student's sentences (the narrow context co-occurence method). As an extension of this method, the key sentence was also matched against the entire set of categories found in any three contiguous sentences of the student's answer (wide context co-occurence method). Since it could be anticipated that the list of synonyms provided for each word of the key might be incomplete, a

technique compromising between the frequency and co-occurrence methods was also programmed. In this method, it was required that at least half the categories of a key sentence be found in the student's response for scoring to proceed; here the student earned a percentage score for the number of categories common to his response and the key divided by the total number of categories in the key sentence, with of course, a minimum non-zero score of 50% (the threshold co-occurrence method). In practice, the threshold method was applied to single sentences (narrow context threshold) and to three contiguous sentences (wide context threshold). Thus five scoring techniques were programmed. The scores for the essays produced by the computer were then correlated with the criterial human grades.

The first finding to note is that the requirement of complete co-occurrence was too stringent; most responses scored this way simply earned zero; in short, it did not permit discrimination necessary for grading. As regards the correlations between judge grade and scores generated by the remaining three methods, the results were quite good (See Table 1).

Insert Table 1 about here

The test of the adequacy of these computer scoring techniques appears to have been limited by the reliability of the human judges. It may be seen, for example, that the computer scores agree as well or better with the pooled judge grades than the judges agree with each other. Most important, the experimental hypothesis was supported by the finding that the threshold, wide context technique for scoring co-occurrence correlated better with the criterion than the word frequency technique. The success of the co-occurrence measures must be interpreted with respect to the kinds of answers required by the particular identification terms forming the student test. Logical development of thought

was not required for a correct answer to any of the identifications, and since the average essay contained only three sentences, there was little opportunity for categories to occur beyond the three sentence unit which had worked so well in this research.

FISHING EXPEDITIONS

In the research projects described above, the flow of activity proceeded from conceptualization of the problem, to the formulation of theory and constructs, to the construction of categories, and ultimately to a test of the empirical relationships deduced from the theory. In this process, the only virtue displayed by the computer was that of counting machine.

However, not all researchers believe that theory is the most useful guide to research. Page, for example, defended the research strategy he employed in the attempt to grade essays by computer by arguing that, "in general, prediction research would be unnecessarily and artificially restrained if it were not permitted use of any convenient predictors, regardless of the vagueness of rationale for their inclusions. There were in this study a fair number of what might be called, therefore, "proxes of opportunity," (Page, and Paulus, 1968, page 25). (Page defines "prox" as any measure which does not provide direct information on a variable of interest, but does approximate it; a list of common spelling errors, if used by the computer to measure errors in text, would be termed a prox).

A fine illustration of the futility of such atheoretical research is that of Dell and Hiller (1971) who attempted to discover correlates of effective lecturing. At the time this study was conducted, dictionaries constructed specifically for analyses of teacher lectures had not yet been key-punched. But the General Inquirer's Harvard Psychosociological Dictionary III was

available and on that basis employed to search for characteristics of teacher effectiveness. Unfortunately, some fish were caught. For example, "Medical terms" demonstrated a correlation of -.56, ($p < .05$) with effectiveness in a lecture about Yugoslavia, and "Sex Themes" a correlation of .75. Having obtained these correlates, the investigators were quite unable to contrive any productive explanations. The computer in such explorations merely absorbs resources better spent on constructive thinking.

Conclusions

The writers conducted a review of the literature pertaining to computer applications in language analysis which failed to uncover any evidence even hinting that use of the computer contributed to the theoretical excellence of research employing it as an instrument of measurement and analysis (with the single possible exception being the research on stylistic analysis reported by Sedelow and Sedelow, 1967). Yet one frequently encounters the argument that use of the computer forces the researcher to make his theory explicit, since programming the computer for analysis absolutely requires a formal, exact statement of the analytic measures. Where this view accurately portrays a body of research, that research may well be described as trivial, (eg. where the theory is stated to be a collection or list of search words). But worse, this view misrepresents the relationship of measurement to theory in permitting the measures (word counts) to go unexplained. In the case of having the computer grade essays by counting the number of commas, hyphens, colons, etc., the research has turned from concern with important characteristics of writing, which the computer cannot measure, to concern for measurement of superficial statistical characteristics of text which the computer

can manage. Furthermore, since no theory is explicated which relates counts based on certain of these statistical aspects of text (hyphens, dashes, slashes, etc.) to the human judgemental process invoked when writing quality is evaluated, it is clear that use of the computer does not insure that theory will be developed and clearly formulated. Use of the computer to spew correlations may thus lead the researcher to overlook significant features of the text to be studied.

It is our opinion that emphasis on the computer as an agent in automated content analysis is misplaced. Lest this opinion appear merely academic, having as its aim a straw-man, we may point to the fact that a new professional journal has appeared — Computer Studies in the Humanities and Verbal Behavior. We see three noteworthy dangers in such emphasis. One, the lay public may be enticed by the scientific appearance of computer research to grant its authors and results unfounded acceptance. Two, considerable resources may be wasted in pursuit of trivial correlations. Three, gradual public realization of some of the severe limitations of computerized content analysis may also produce an unjustified rejection of this form of research.

References

Carlson, A.R., Concept frequency in political text: an application of a total indexing method of automated content analysis. Behavioral Science, 1967, 12, 68-72.

Gage, N.L., Belgard, M., Dell, D., Hiller, J., Rosenshine, B., and Unruh, W., "Explorations in the teacher's effectiveness in lecturing." In I. Westbury and A. Bellack, Research into Classroom Processes. New York: Teacher's College Press, 1971, 175-224.

George, A.L., "Quantitative and qualitative approaches to content analysis." In Pool (Ed.), Trends in Content Analysis. Urbaba: University of Illinois Press, 1959, p. 244.

Hiller, J.H., Verbal response cues to conceptual vagueness. Paper read at the meeting of the American Educational Research Association, San Francisco, February, 1969.

Hiller, J. H., Fisher, G.A., and Kaess, W. A computer investigation of verbal characteristic of effective classroom lecturing. American Educational Research Journal, 1969, November, 661-675/

Hiller, J.H., Marcotte, D., and Martin, T. Opinionation, Vagueness, and Specificity-Distinctions: essay traits measured by computer. American Educational Research Journal, 1969, March, 271-286.

Holsti, O.R., "A Computer Content-Analysis Program for Analyzing Attitudes: The Measurement of Qualities and Performance." In Gerbner, Holsti, Krippendorff, Paisley, and Stone (Ed.), The Analysis of Communication Content. New York: Wiley, 1969. 381-405.

Iker, H.P., and Harway, N.I., "A Computer Systems Approach Toward the Recognition and Analysis of Content." In Gerbner, et. al., (Ed.), The Analysis of Communication Content. New York: Wiley, 1969. 381-405.

Laffal, J., Pathological and Normal Language. New York: Atherton Press, 1965. p. 249.

Mahl, G.F., "Exploring emotional states by content analysis." In Pool (Ed.), Trends in Content Analysis. Urbana: University of Illinois Press, 1959. p. 244.

Marcotte, D.R., "A computerized contingency analysis of content graded essays." Unpublished doctoral dissertation submitted to the University of Connecticut, 1969.

Sedelow, S.Y., and Sedelow, W.A. Jr., "Stylistic Analysis." In Borko, H. (Ed.), Automated Language Processing. New York: Wiley, 1967. p. 386.

Stone, P.J., "Improved quality of content-analysis categories: computerized disambiguation rules for high frequency English words." In Gerbner, et. al. (Ed.), The Analysis of Communication Content. New York: Wiley, 1969. 199-223.

Stone, P.J., Dunphey, D.C., Smith, M.S., and Ogilvie, D.M. The General Inquirer: A Computer Approach to Content Analysis. Cambridge: M. I. T. Press, 1966, p. 651.

FOOTNOTE

[1]An earlier version of this paper was presented at the Association for Computing Machinery National Conference, San Francisco, Aug. 1969.

Category	Number of Category Items
Ambiguous designation (all of this, and things, somewhere, other people	61
Negated intensifiers (not all, not many, not very)	57
Approximation (about as, almost, pretty much)	29
"Bluffing" and recovery (a long story short, anyway, as you all know, of course)	55
Error admission (excuse me, not sure, maybe, I made an error)	18
Indeterminate quantification (a bunch, a couple, few, some)	28
Multiplicity (aspects, factors, sorts, kinds)	36
Possibility (may, might, chances are, could be)	17
Probability (probably, sometimes, ordinarily, often, frequently)	19
Reservations (apparently, somewhat, seems, tends)	33
TOTAL	353

FIG. 1. VAGUENESS CATEGORIES. (The latest vagueness dictionary has been modified to include common pronouns, eg. this, that, it, etc.).

Table I

Correlation between judge grades and computer scores for 963 essays.

	Median	Range
Frequency	.495	.30 to .61
Threshold narrow	.530	.21 to .71
Threshold wide	.605	.33 to .74
Average correlation of each of 8 judges with the 7 others*	.475	.36 to .53

* Fisher Z transforms were applied prior to averaging.